Multimedia Imaging Technology

THE UNIVERSITY OF AUCKLAND
FACULTY OF SCIENCE

Approximated Ground Truth for Stereo and Motion Analysis on Real-World Sequences

Zhifeng Liu and Reinhard Klette

Outline and Objective

Abstract

We approximate ground truth for real-world stereo sequences and demonstrate its use for the performance analysis of a few selected stereo matching and optic flow techniques. Basically we assume zero roll and constant tilt of an ego-vehicle (for about 10 seconds) driving on a planar road.

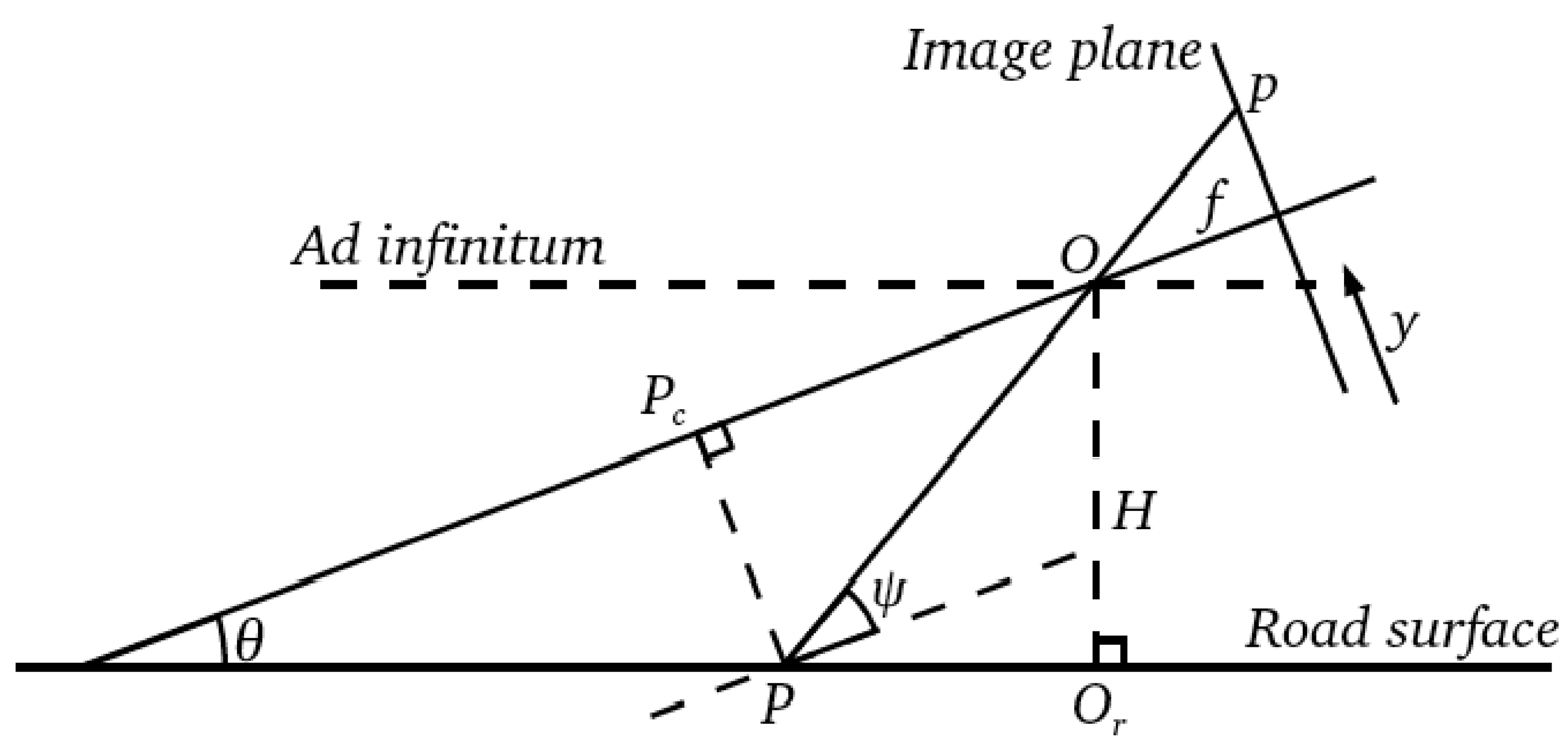
Our Approach

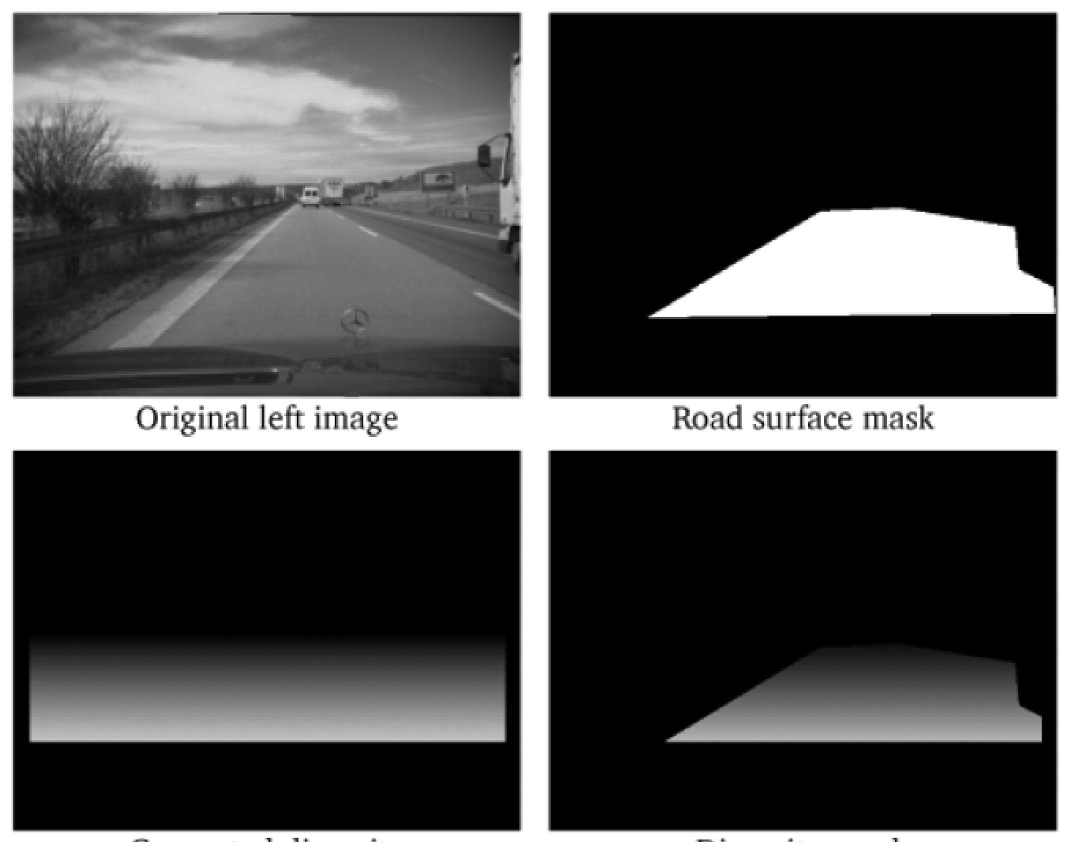
Similar to [Labayrade et al. 2002], vertical disparities are used to approximate the road surface by planar patches, say every 10 frames in a sequence. This allows to approximate ground truth for stereo and optic flow `on the road', and to evaluate low-vision algorithm results on those areas.

Approach

Generation of Road Mask for Stereo Ground Truth

We use a few (`secure') disparity values to estimate the road surface in front of the ego-vehicle as part of a plane. Knowing pixel position *p*, focal length *f*, height *H*, angle ψ, and disparity *d* at pixel position *p* (for3D point *P*), we can calculate tilt angle θ. The focal point of one of the stereo cameras is calibrated to be at *O*. This allows to generate a disparity mask for a segmented road area.





The table below shows estimated tilt-angles for a stereo sequence of 240 stereo pairs, estimated after every 10 stero pairs.

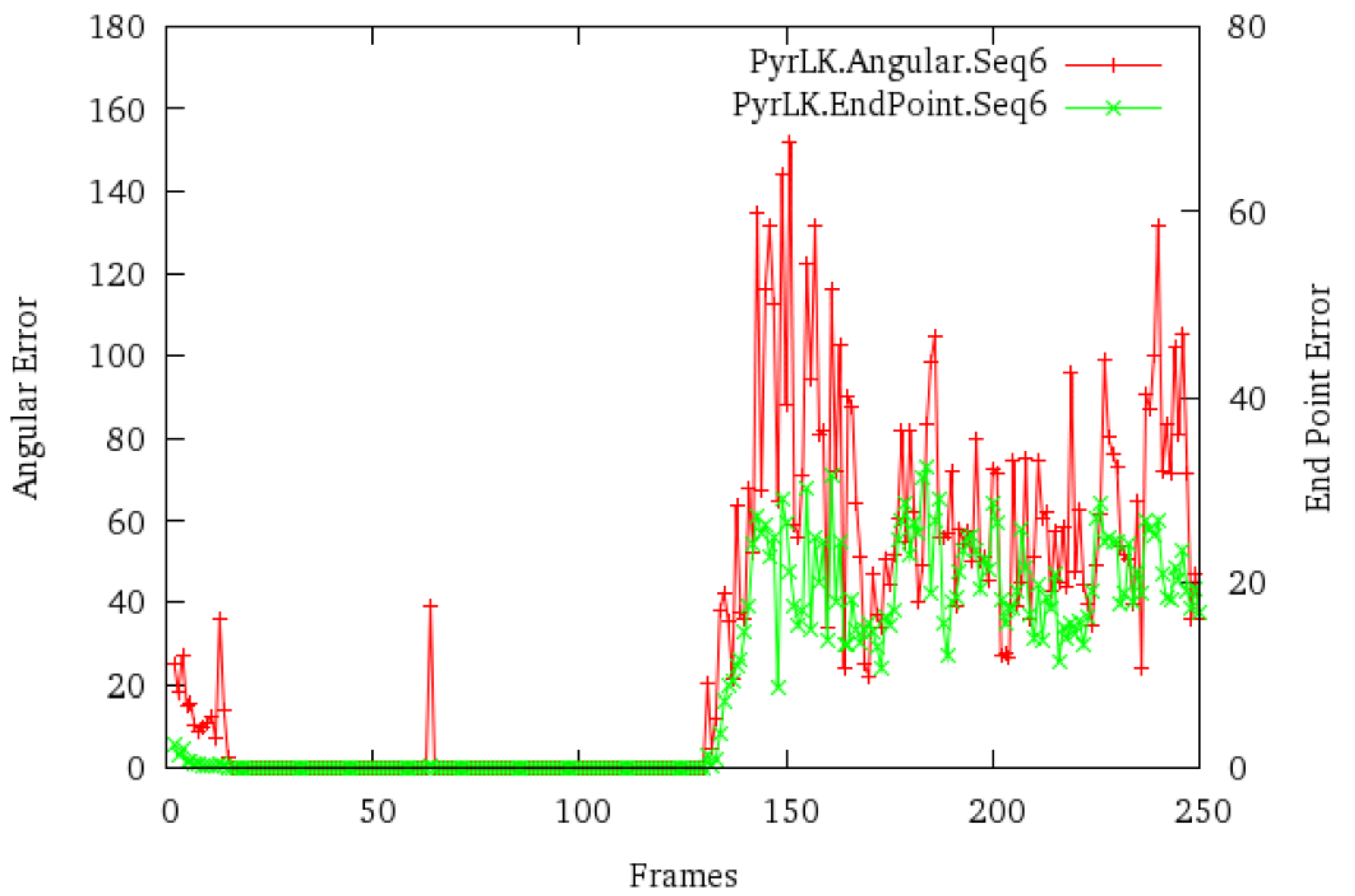
First pair of frames	1	11	21	31	41	51	61	71	81	91	101	111
Tilt angle (10^{-3} of a radian)	80	71	60	60	62	63	65	70	77	71	63	66
First pair of frames	121	131	141	151	161	171	181	191	201	211	221	231
Tilt angle (10^{-3} of a radian)	60	50	50	59	58	54	55	56	58	53	53	42

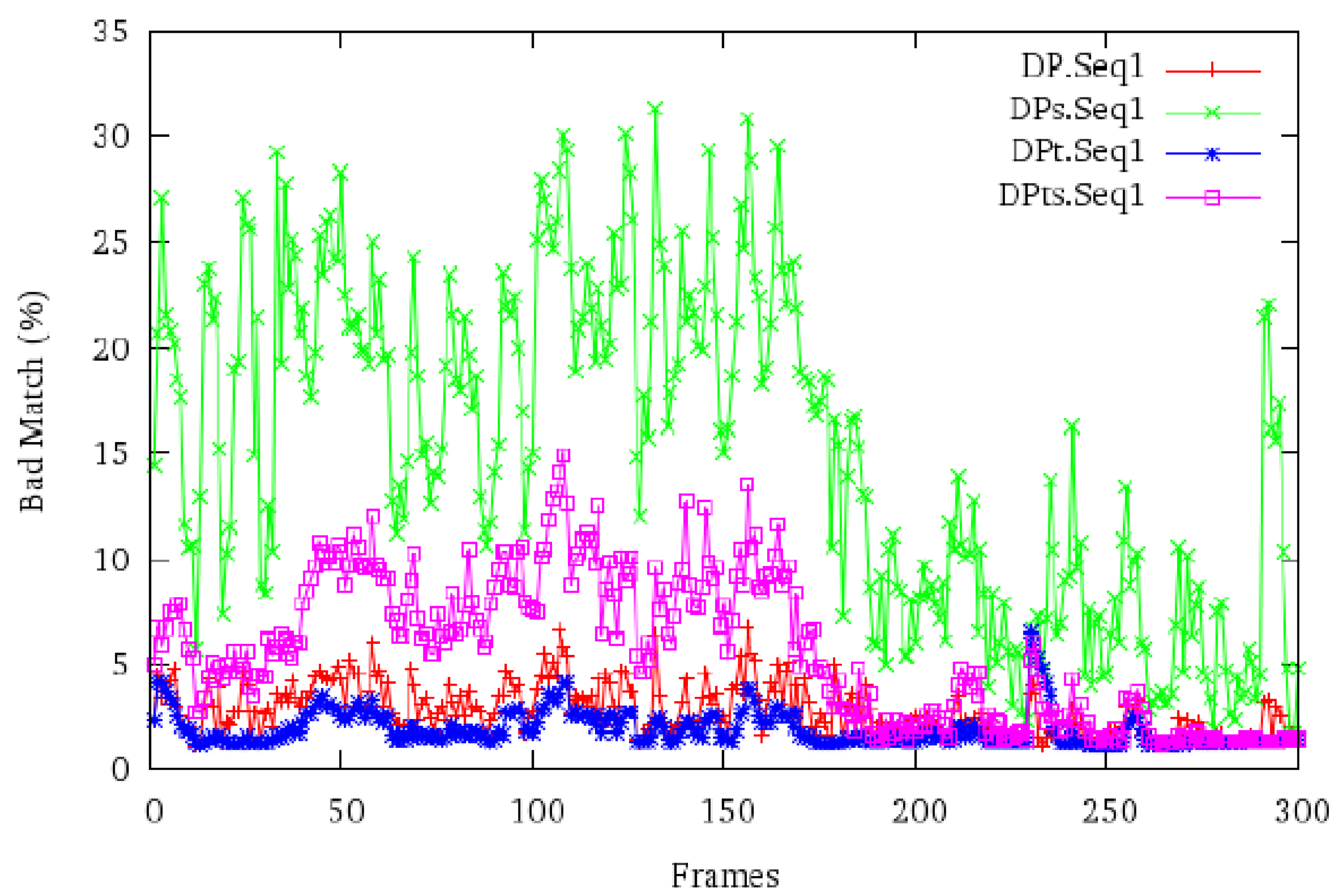
Generation of Local Disparity Ground Truth on Road Surface

Applying similar elementary geometry, the local displace–ment of one road surface point, from frame at time t to time t+1, may be calculated. (LSE optimization is used in both cases).

Results

Example (right): percentages of bad matches for different versions of dynamic programming stereo on a sequence having 300 stereo pairs. It points out that temporal propagation is rated best (on the road surface!), and this is not in consistence with results for engineered Middlebury stereo images where spatial propagation (Birchfield-Tomasi) is rated best for dynamic programming.





Example (left): mean angular and end point errors for each frame, in a sequence of 250 frames (pyramidal Lukas-Kanade). Sequence analysis allows a new quality of performance evaluation, compared to very short sequences only.

The order of the algorithms' performance is clearly inconsistent to that reported on the Middlebury stereo or optical flow website. This difference shows the necessity for establishing performance evaluation methods also (!) on various real-world sequences.